Fundamentals of Multimedia Computing

Chapter 4: Light

Authors:
Gerald Friedland
and
Ramesh Jain

Draft for Comments

Light

Light is one of the most basic phenomenons in universe. The Bible begins with the words “Let there be light!” As a consequence most of our brain is dedicated to understand the reflection of light from objects that hit our retina to form an image of our surroundings. Many of mankind’s innovations have evolved around humans being able to capture that image and store it. First the painters in the caves in stone age, then the improved sculpturing and painting in the antique being followed by middle age and renaissance painters. Finally, photography, filming and then digital storage of movies and photographs. Most recently, a computer science discipline evolved around computer-based interpretation of images, called computer vision. In this chapter, we are going to introduce the basic properties of light and how it is stored and reproduced. Basic image processing and introductory computer vision techniques will be discussed in later chapters.

What is light?

In contrast to sound, which is clearly defined as a wave with a certain frequency traveling through matter, physicists know that light has both wave as well as particle properties. It is beyond the scope of this book to discuss the nature of light in depth. We will define light therefore as the portion of electromagnetic radiation that is visible to the human eye. Visible light has a wavelength of about 400-780 nano meters, which corresponds to a frequency of 405-790 THz. The adjacent frequencies of infrared on the lower end and ultraviolet on the higher end are still called light, even though there are not visible to the human eye. The traveling speeds of light in a vacuum is one of the fundamental constants of nature as it is the fastest speed observable, and is 299,792,458 meters per second. Apart from speed, primary measurable properties of light are intensity, propagation direction, polarization, phase, and, as discussed, frequency or wavelength. The phase is the fraction of a wave cycle which has elapsed relative to an arbitrary point and can be manipulated by filters to change the appearance of light. The polarization describes the orientation of of the light waves. All electromagnetic waves, including light and gravitational waves may exhibit polarization. Sound waves in a gas or liquid do not have polarization because the direction of vibration and direction of propagation are the same. The orientation of the electric fields produced by the light emitters may also not be correlated, in which case the light is said to be unpolarized. However, if there is at least partial correlation between the emitters, the light is said to be partially polarized. One may then describe the light in terms of the degree of polarization. It is possible to build filters that only allow light of a certain degree and angle of polarization, an effect that is often used in 3D vision: 3D glasses often have filters for the left and the right eye that allow for different polarized light to go through for each eye and therefore allow the two eyes to see images with slightly different disparity.

The intensity of light is measured in three different units: candela, lumen and lux. The candela (cd) measures of luminous intensity, which is defined as power emitted by a light source in a particular direction, weighted by the luminosity function -- a standardized model of the sensitivity of the human eye to different wavelengths, see Figure 1 and references. The unit is originally derived from the light emitted by a common candle (thus its name). A standard candle pretty much emits light with a luminous intensity of roughly one candela. The physical definition is as follows: The candela is the luminous intensity, in a given direction, of a source that emits

monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of $\frac{1}{683}$ watt per steradian. A 100 W incandescent lightbulb emits about 120 cd.

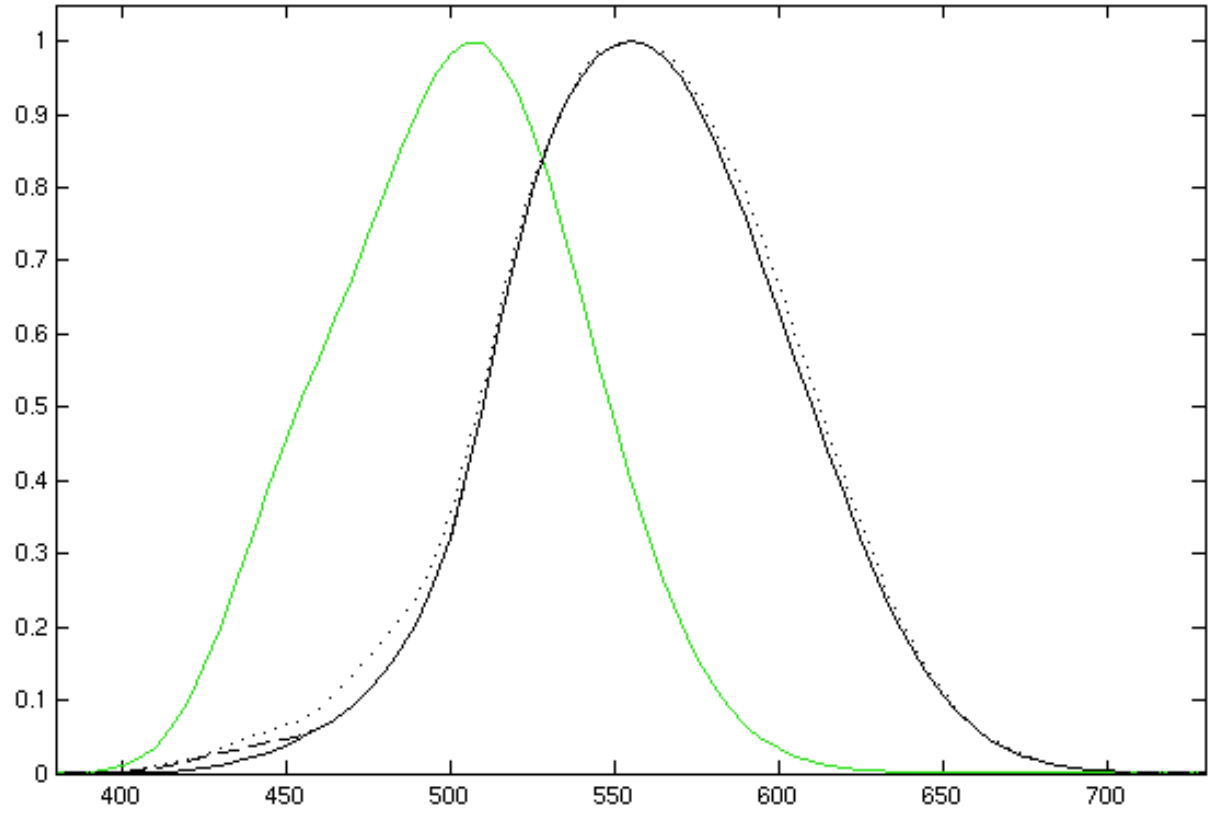

Figure 1. This graph shows different luminosity functions describing the sensitivity of the human eye to light of different wavelengths. Several luminosity functions are currently in use (see references). The dotted is the most currently accepted luminosity function from 2005.

The lumen (symbolic abbreviation lm) is unit of luminous flux. The lumen is defined in relation to the candela as

$$1\ lm = 1\ cd{\cdot}sr$$

As a full sphere has a solid angle of $4{\cdot}\pi$ steradians, a light source that uniformly radiates one candela in all directions has a total luminous flux of 1 cd·4π sr = 4π ≈ 12.57 lumens. The light output of video projectors is typically measured in lumens. The America National Standards Institute (ANSI) standardized a procedure for the measurement of the light output of video projectors, which is why many projectors are currently sold as having a certain amount of "ANSI lumens" even though ANSI did not redefine lumen as a physical unit.

Lux (symbol: lx) is the physical unit of illuminance and luminous emittance measuring luminous power per area. The unit is equivalent to watts per m^2 (power per area) but with the power at

each wavelength weighted by the luminosity function (see Figure 1). Lux, Lumen and Candela can be converted into each other as the following equation holds:

$$1\ \mathrm{lx} = 1\ \mathrm{lm/m^2} = 1\ \mathrm{cd \cdot sr \cdot m{-}2}.$$

A full moon overhead at tropical latitudes is said to emit about 1 lx of light. Office lighting is usually at 320-500 lx, TV studio lighting is at 1000 lx. Inside the visible frequency range, the human is able to see as little as one photon in the dark and yet the eyes can be opened in a dessert at noon with sun exerting up to 130,000 lx. This is an incredible adjustment that human-made light sensors are currently rarely capable of.

Observed Properties of Light

Similar to sound light exhibits properties while traveling through space. Like sound light is mostly not exclusively traveling in a homogenous medium from a source to exhaustion. Especially, for recording lenses are typically used (see later). Also, the environment is filled with objects that may absorb, dampen, or reflect light. Sometimes, especially outside, other light sources may appear and collide with light waves in question. Again, the resulting effects of these conditions play a large role when designing multimedia systems. However, for practical purposes light waves are much less effected by environmental conditions than sound waves.

Reflection of light is simply the bouncing of light waves from a certain object back into a similar direction the light was coming from. Often, energy is absorbed (and converted into heat or other forms) when reflecting from an object so the reflected light may have slightly different properties as it may have lost intensity, shifted in frequency, polarization, and so on. Light is usually absorbed by solid objects, such as a concrete wall. Meaning, light waves cannot travel through them. The opposite property is called transparency: When light can ravel through an object seemingly unchanged. the object is called transparent. Detecting transparent objects is probably one of the most challenging tasks in vision, including computer vision.

The most important effect observed when light passes through a transparent object is called refraction. Refraction is the “bending” of light rays when passing through a surface between one transparent material and another. When a beam of light crosses the boundary between two different media (including vacuum), the wavelength of the light changes, but the frequency remains constant. If the beam of light is not crossing the boundary in an orthogonal angle, the change in wavelength results in a change in the direction of the beam. This change of direction is known as refraction and can be observed in every day item, for example when trying to grab a fish in an aquarium or observing a “bending” straw in a glass of water. The study of light and the interaction of light and matter is termed optics. Since this is a book on multimedia computing we cannot exhaustively discuss all properties of light here and optics is its own field of study. However, a common phenomena sometimes neglected but also sometimes hated by multimedia researches is that of lens distortion. In a lens, a uniform light beam impacts the transparent object with varying angles, so the the refraction also varies. Therefore, the image that is projected through the lens is distorted. Figure 2 shows examples of typical distortions.

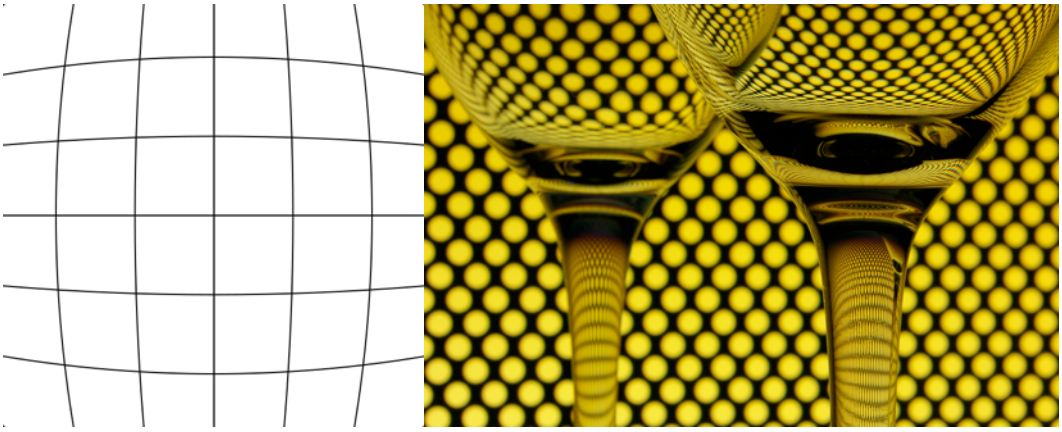

Figure 2. Left: Typical lens distortion pattern, right: a picture from Wikimedia Commons showing the distortion created by wine glasses (picture by Atoma, creative commons license, needs to be contacted)

Correcting lens distortion can be a serious issue. When possible, distortion can be corrected by calibration, i.e. by projecting a well-defined object onto the lens, e.g. a grid a shown in Figure 1, and calculating a correction function between the actual image and the distorted image. In many cases, however, only the projected image is given and distortion is hard to correct.

Recording of Light

Light can be stored and reproduced as images and video. The device to this is called a camera. The term camera is derived from the latin word camera obscura (“dark chamber”), an early mechanism for projecting but unable to store images. Figure 3 shows historical sketches of the mechanism. The device consists of a box (that can be as the size of a room) with a hole in one side. Light from an external scene passes through the hole and strikes a surface inside where it is reproduced, upside-down, but with both color and perspective preserved. The modern camera evolved from the camera obscura: The image is projected onto a light sensitive memory, which in the beginning was light sensitive chemical plates, later chemical film and nowadays is photosensitive electronics which can record images in a digital format.

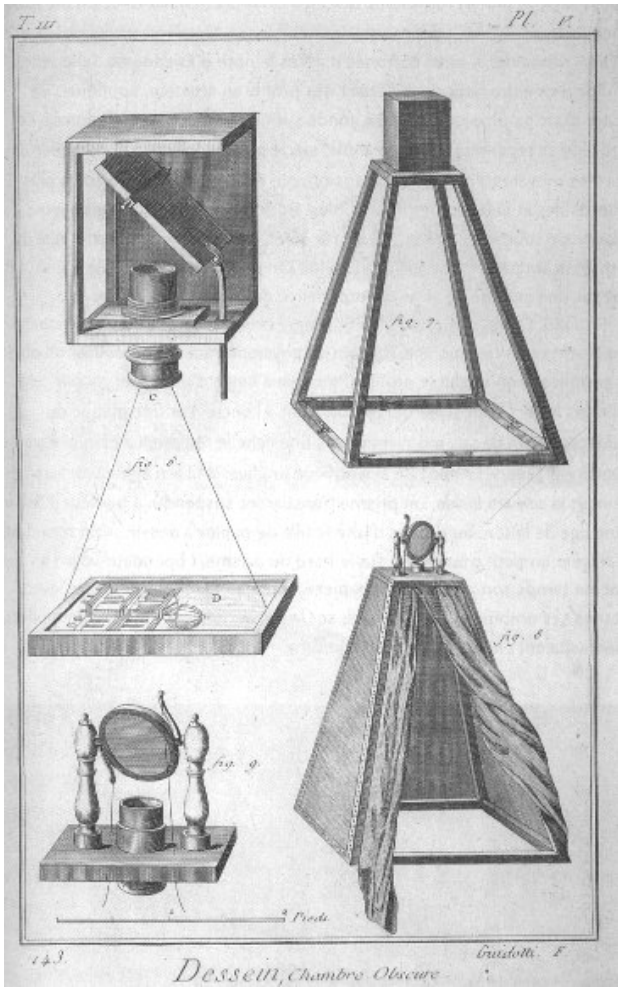

Figure 3. Historical drawings of a camera obscura often used for paintings.

The early photographic plates consisted of a glass plate covered with light-sensitive emulsions of silver salts. The salts turned light when exposed to light, leaving a gray-toned photograph behind. Photographic plates largely faded from consumer market in the beginning of the 20th century when the more convenient and less fragile films were introduced. However, the plates were still used for some scientific photography until digital photography arrived because the professional astronomical community had very good use for material that responds to very little light, especially in large-format frames for wide-field imaging. Through a chemical process the negative images were converted to positive images which are easier to interpret to the human eye, i.e. light impact equals dark color.

Photographic film is a sheet of plastic coated with an emulsion of light-sensitive silver halide salts with variable crystal sizes that determine the sensitivity, contrast, and resolution of the film. Note that contrary to the popular believe that one can zoom in arbitrarily into analog film because there is no pixels, analog pictures do have a maximum resolution albeit usually much higher than current digital images. When the emulsion is sufficiently (i.e. light with enough intensity for a long enough time) exposed to light, or other forms of electromagnetic radiation such as X-rays, it forms an invisible image. Chemical processes can then be applied to the film to

create a visible image. This process is called film developing. In black-and-white photographic film there is usually one layer of silver salts. Color film uses at least three layers. Todays films usually many more, e.g. up to twenty. Dyes, absorbing to the surface of the silver salts, make the crystals sensitive to different colors. A very important property of analog films is the so-called film speed. Despite it's name, film speed is not a velocity but describes a film's sensitivity to light exposure. There are different standardized film speeds but the most commonly known is the ISO rating. Consumer-rated films are usually labelled with ISO 100, 200, 400, or 800, where a lower number determines longer times the film must be exposed to light for a proper photograph. The speed is determined by a ratio of the optical density of the material and the logarithm of the exposure time. The logarithm is taken because film material usually reacts non-linearly to light exposure -- which is why many professional digital cinematographic multimedia file formats still use a logarithmic sampling scale. Recording movies on a chemical film usually requires taking 25 images or more per second. Of course, a film with the appropriate film speed has to be chosen which makes recording of movies in darker light more difficult and this is way cinematographic filming generally requires more sophisticated lighting then photography.

Nowadays, most photos are recorded electronically. A digital camera still follows the original principle of the Camera Obscura but instead of a chemical reaction on a plate or a film, a physical reaction occurs in an electrical photovoltaic element, typically a so-called CCD sensor chip (Charge-coupled device). So in other words, a modern camera is a visual sensor that converts light into an electrical signal. As with regular films, there is a maximum granularity, which in digital cameras is defined by the number of photoelectric sensors. Each sensor creates one picture element (also known as pixel). The number of pixels is usually given as the maximum granularity of a picture, which in the digital world is called resolution. Typical photo cameras have a resolution of several megapixel -- millions of pixel. Resolutions of the resulting image are usually specified as X*x*Y axis resolution, e.g. 1024x768 or 1280x1024. While this might change in the future, todays' digital cameras don't have the memory to store images by representing each pixel directly as a sensor value. Therefore, images are usually compressed by applying spectral compression (see Chapter XXX). A typical image formats that uses this type of compression is the JPEG format. The format is described in later chapters. Uncompressed images are rare but sometimes needed for content analysis (see Chapter XXX) and for editing of high-quality images -- these are called raw images.

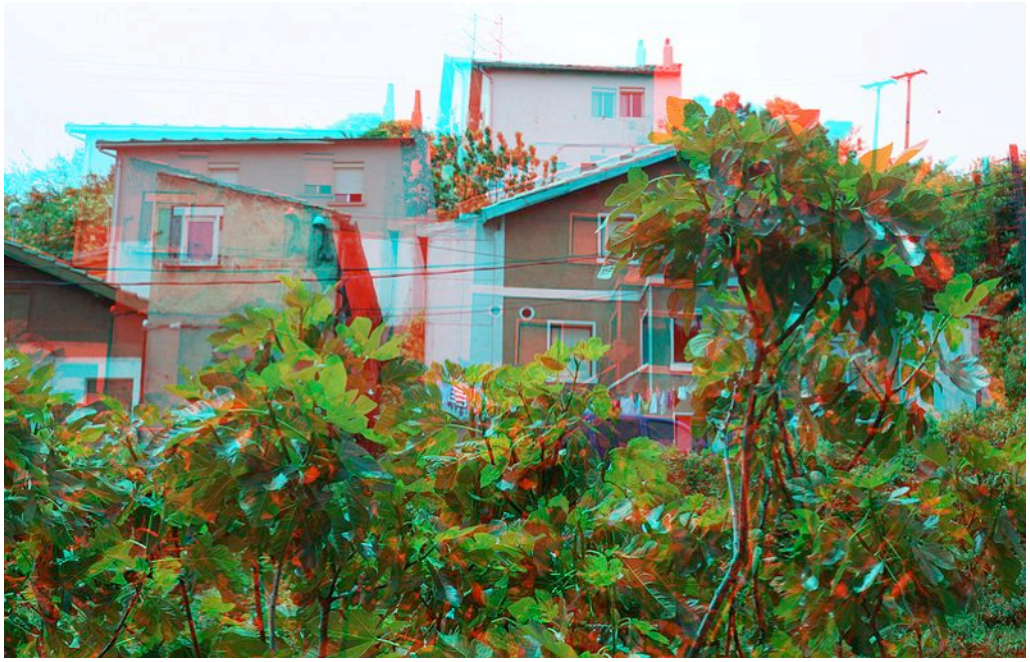

Figure 4. Anaglyph 3D photograph viewable with red/green glasses. If you are viewing this photo on a computer screen and the 3D effect does not work, adjust your display settings to match the filters in your glasses. THIS Photo must be PRINTED in COLOR!

Video cameras have recorded light in electrical manner for a longer time. TV cameras have evolved as analog devices storing the electrical changes on the CCD sensors on magnetic tapes and transmitting them through the air using analog radio waves. Video cameras also record sound at the same time, which is described in Chapter XXX. It's important to realize, that for many years, cinematic cameras where still using chemical film (usually the so-called 35mm film) because TV cameras only delivered images with very small resolutions not suitable for "the big screen". Typical TV resolution were PAL, SECAM (720x576 analog picture elements), and NTSC (640x486 analog picture elements). The formats' color encodings are very different and will be discussed later in this chapter. The introduction of digital video cameras not only made the photographic and the videographic world converge, it also allowed videos to be recorded in a much higher resolution, especially because image compression methods could be modified to support moving pictures. Modern cameras store videos directly in compressed format, such as MPEG (see Chapter XXX). The resolution of digital photo and video cameras increase constantly. At the time of writing this book photo cameras with up to 32 megapixel and videocameras with up to 16 megapixel exist on the market.

While photographic and cinematographic recordings as described on the preceding pages can only be performed as a projection onto a surface, this does not mean that the resulting images have to be flat, i.e. 2-dimensional. Of course, actual reflection in space is 3-dimensional and humans are able to perceive the distances between objects in space 3-dimensionally. While we will talk about visual perception later, the desire to capture scenes with depth is relatively old and the first commercial 3D photo cameras date back to 1947. For capturing 3D images, the predominant technology is the stereo camera. A stereo camera has two (or more) lenses and a

separate photographic sensor (of film) for each length. This allows the camera to simulate human two-eyed vision, which is the basis for depth perception. The distance between the lenses in a typical stereo camera (the so-called intra-axial distance) is usually chosen to be about the distance between a human's eyes (known as the intra-ocular distance), which is about 6.35 cm. However, a greater inter-camera distance can produce pictures where the 3-dimensionality is perceived as more extreme. This technique works with both images as well as movies, provided that the images are kept separate and only one eye is exposed to each image. Therefore, for watching a 3D movie, usually polarizing or red/cyan filter glasses (so-called anaglyph technique) have to be worn. These separate the two images even though shown by the same projector. The two images are superimposed through two filters, one red and one cyan, or two polarizing filters. Glasses with colored/polarizing filters in each eye separate the appropriate images by canceling the filter color/polarization out and rendering the complementary color/polarization black. While other technologies exist to create 3D projection, including auto-stereoscopic methods that do not require glasses, these two techniques are, as of the time of writing this chapter, predominant. Figure 4 shows an example of a 3D photography.

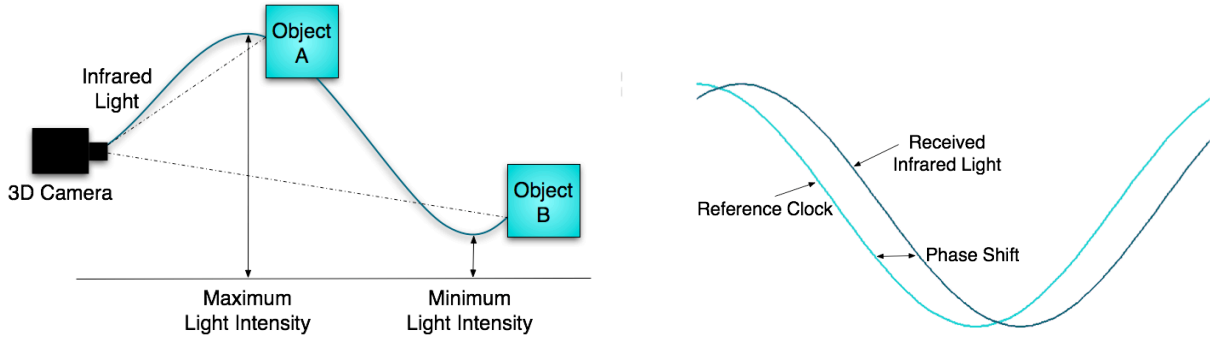

Figure 5: Left image: Two objects reflect amplitude-modulated infrared light. Object A reflects more light than object B because at the point of time when the photons hit object A, they were emitted with maximum light intensity. The photons that hit object B at the same time where emitted before that, with lower intensity. Right image: The actual distance can be calculated by measuring the phase shift between the emitted and the reflected light. If the distance of the reflecting object were zero, the two curves would have no phase shift. The farther the object away, the greater the phase shift.

Most importantly, these techniques aim at human perception and require the human brain to "decode" the stereoscopic image. Computing the depth encoded in a stereoscopic image is still a matter of research (compare references). Therefore, different types of devices are currently in research that try to estimate depth information in a way that it is directly available to a computer. One of these technologies is the so-called time-of-flight camera. A time-of-flight camera works very similar to radar. The camera consists of an amplitude-modulated infrared light source and a sensor field that measures the intensity of backscattered infrared light. The infrared source is constantly emitting light that varies sinusoidal. Object A reflects almost the maximum intensity while object B, having a greater distance to the camera, reflects less light. This is because at any specific moment, objects that have different camera distances are reached by different parts of

the sinus wave. As shown in Figure 5, the incoming light is then compared to the sinusoidal reference signal which triggers the outgoing infrared light. The phase shift of the outgoing versus the incoming sinus wave is then proportional to the time of flight of the light reflected by a distant object. This means, by measuring the intensity of the incoming light, the phase-shift can be calculated and the cameras are able to determine the distance of a remote object that reflects infrared light. The output of the cameras consists of depth images and a conventional low-resolution gray-scale video, as a byproduct. While the idea is promising, current technological realizations are still facing problems with artifacts caused by quickly moving objects, light scattering, background illumination, or the non-linearity of the measurement. Last but not least, time-of-flight cameras still require a large budget compared to regular cameras. Also, the reproduction of a time-of-flight recording in 3D is not straightforward.

Reproduction of Light

When trying to reproduce a specific light pattern, there are mainly to methods: additive and subtractive. Subtractive methods rely on intensity variations of the reflection of ambient light and do not work when no light is present. The simplest example is paper. Paper reflects patterns differently once it has been modified by ink or toner, which can result in visible patterns. Additive methods work with active light sources that are mixed together, the most common example is CRT display in a TV, as explained below.

Photographic plates and film relied on light reflection for the reproduction of light, sometimes helped by a projector with a powerful light bulb, e.g. for movies or transparency shows to a larger audience. However, electrical recording of light allows for an active reproduction using light sources. As mentioned earlier, the first electric storage and transmission of light was deon using TV equipment. As a result, the first technology for reproducing (moving) images electrically was the TV. Although, the earliest TVs were made in the late 1920s, TV really took off after World War II in the 1940s. The technology adapted by these TVs was the cathode ray tube (CRT), that was invented by German Telefunken company in 1934. A cathode ray tube is a vacuum tube with a source of electrons that are projected onto a fluorescent screen. The fluorescent material on the screen reflects light when hit by the electron beam. The beam is controlled by an electromagnetic field that accelerates and/or deflects the beam in order to control its impact on the fluorescent surface, thereby controlling the amount of reflection -- forming a grayscale image. To produce color TV, which was not introduced in many countries before the 1970s, a CRT with three different phosphors which emit red, green, and blue light respectively is used. The reflective phosphors are packed in clusters called triads. Roughly speaking, one triad corresponds to one color pixel. The colors red, green, blue where chosen for technical reasons since they are sufficient to mix most perceivable colors by using different strengths of red, green, and blue. The use of red, green, and blue (RGB) triads was so predominant that modern graphic cards and displays are still using it, even though the eye's perception of light uses different base frequencies (see later in this chapter). One important characteristic of CRT triodes is called "gamma", this is a non-linear function between applied voltage of the electron gun and light intensity in the reflection. The non-linearity of the image response often results in artifacts that are perceived as image distortions, especially in dark images. Therefore, often a so-called gamma-correction function is applied to help normalizing the perceived image. Two other important characteristics of a CRT are its vertical and horizontal

frequencies which are describing the frequency the beam is visiting each line and each column of the display. In analog TV vertical frequencies were usually adjusted to fit the frequency of the power outlets therefore US NTSC uses a 60Hz base frequency and European-based PAL and SECAM use 50Hz. In order to double the perceived frame rate, analog TV uses interlacing. It allows fast motions to appear unjittered and works by only updating every second line of screen each frame. First the odd lines then the even lines then the odd lines again, and so on. As a side effect, the signal bandwidth can be halved since only have of the information has to be transferred per frame. While the advantages outweighed the disadvantages in analog TV, the disadvantages become clearly visible in comparison o high-resolution digital TV. Also, digital video compression and content analysis algorithms often have issues with interlaced video. Figure 6 shows an example of typical interlacing artifacts, the so-called interline Twitter. Modern digital video encoders usually have a feature to de-interlace image frames. Since the lost information cannot be precisely restored, however, de-interlacing algorithms use heuristics to guess the content of the lost lines. A common method is to duplicate lines or to interpolate between two lines.

Figure 6. Two interlaced video frames showing a fast motion.

Modern TVs and computer monitors do not use CRT anymore. They use different technologies that allow higher resolution in update time and image granularity, save energy, and are less bulky than CRT displays and therefore allow overall larger screens. Plasma displays, which have been common for a number of years, rely on the pixel to be displayed by plasma cells. These are in essence gas-filled fluorescent lamps. A display typically consists of millions of plasma cells compartmentalized in between two panels of glass. The cells hold a mixture of noble gases and a small amount of vaporized mercury. Vaporized mercury in combination to an applied voltage makes the gas in the cell become plasma and when the electrons flow through cell striking the mercury, the energy level is raised with excess energy being converted to ultraviolet light. The ultraviolet light is the reflected by a phosphor-painted reflective area on the back of the cell converting the UV light into visible light. Depending on the phosphors used, different frequency ranges of visible light can be achieved, resulting in different colors. Like in a CRT display, each pixel in a plasma display is made up of a triad so that varying the voltage of the signals to the

cells allows different perceived colors. While excellent for TV, the phosphoric reflective layer in plasmas and CRTs should not be used to display the same image for too long as it might damage the phosphoric layer permanently and cause so-called image burn-in. This was the actual reason for people to invent screen savers, which would make sure the image changes constantly. Since about 2008 both CRT and plasma displays have lost in attraction on the market, giving way to the even more lightweight Liquid Crystal Displays (LCDs). While inferior to plasma in the beginning of the development, LCDs now dominate the display market. Nowadays, LCDs are usually more compact, lightweight, less expensive, and more reliable than CRT or plasma displays. They are available in a wider range of screen sizes, and since they do not have to use phosphor as reflective layer (or have no reflective layer), they cannot suffer image burn-in. LCDs are more energy efficient and offer safer disposal than CRTs. As the name implies, LCDs base on liquid crystals. These do no emit light but are able to modulate light, i.e. change the polarization of a light wave. Each pixel of an LCD typically consists of a layer of molecules aligned between two transparent electrodes, and two perpendicular polarizing filters. The electrodes are in contact with the liquid crystals and can align the crystals in a particular direction. If there was no liquid crystal between the filters, light passing through the first filter would be blocked by the second, perpendicular, polarizing filter and appear black. Now with the liquid crystals in place and without a voltage applied to the electrodes, the crystals are unaligned and modulate the light in random direction making the display appear gray. With a voltage applied the crystals align according to the current and modulate the light more and more in a particular direction. With enough voltage applied, the display appears black again. Varying the voltage therefore varies the amount of light passing through the filters which is perceived as varying shades of gray. Since liquid crystals do not emit light themselves, gray scale displays have a reflective layer behind the second polarizing filter to reflect incidental light. Color displays use a light source instead that varies in colors, usually an RGB triad is used. The inexpensive availability of LC displays enabled TVs to become 60” in diagonal and larger, which also prompted demand for higher resolution in TV helping the HDTV standard (which had been invented many years ago) to become popular. Full HDTV is now at a resolution of 1920x1280 pixels and 120Hz refresh rate.

The reproduction of 3D photos and videos has already been discussed in the previous section. The major challenge is to create displays that do not require special viewing devices, such as glasses. Devices that achieve this are called autostereoscopic. Nintendo’s portable game console 3DS is implementing an autostereoscopic display using the so-called parallax barrier method. The parallax barrier is placed in front of the LCD. It consists of a layer of material with a series of precision-angle slits, guiding each eye to see different set of pixels based on the angular direction of focus for each eye. A disadvantage of the technology is that the viewer must be positioned in a well defined spot to experience the 3D effect. Therefore it is currently mostly used for small displays, such as portable gaming systems.

Perception of Light

While it took humankind a while to invent methods for recording and storing light, mechanisms for doing that already existed for much longer. Nature had invented the eye connected to a brain already in very early organisms. As we will see further in this book, multimedia computing cannot be understood without at least a basic comprehension for how human vision works. In fact, the more we learn about how human vision works, the more we can make computer systems and algorithms adapt to it and thereby increase the (perceived) performance of them.

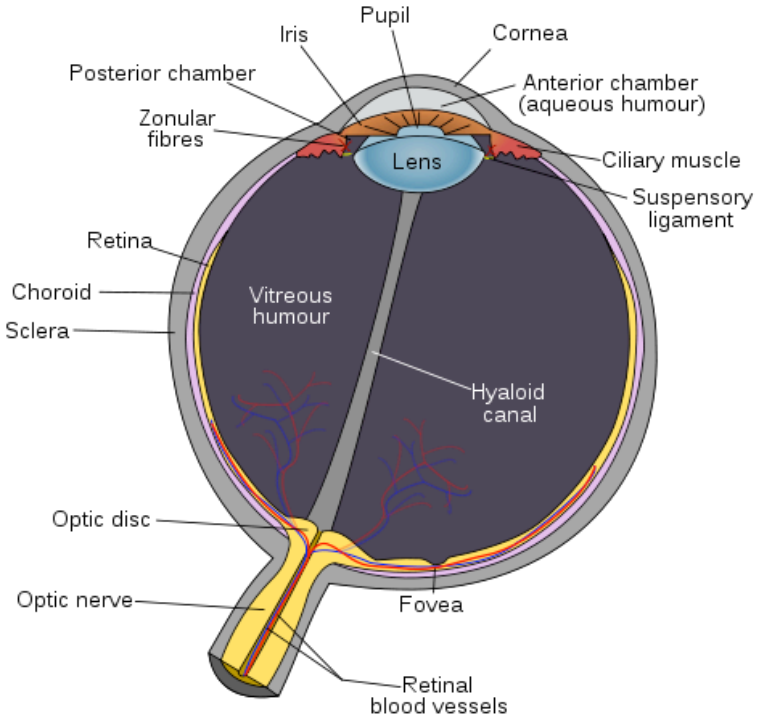

Figure 7. Schematic diagram of the human eye. The rods and cones described in this chapter are found on the retina.

Like a digital camera, an eye detects light and converts into electrical impulses. Figure 7 shows a schematic image of a vertebrate eye. In most higher organisms the eye is a complex optical system that collects light from the surrounding environment, regulates its intensity through a diaphragm, focuses on certain points through an adjustable assembly of lenses to form an image, converts this image into a set of electrical signals, and finally transmits these signals to the so-called visual cortex of the brain. So in many aspects an eye works is a very complex camera obscura. We cannot explain all the processes involved in human vision because this would fill several books and most importantly, human vision is not completely understood. However, in the past, multimedia computing has used several important properties of human vision to an advantage. These are discussed quickly as follows.

One of the most important properties of the human eye is the fact that images shown in a fast-enough sequence are blurred together and perceived as one, enabling video. This property is present in all animals in a certain way, however, different eyes may have different frequency

thresholds. Human's threshold is at about 20-25Hz to perceive objects as a movie rather than (flickering) still images, as explained before 25-30 images per second is the framerate most video technologies work with (as a note: compare this to the lowest frequency acoustic stimulation is perceived as tone rather than period beats).

Like most sensory organs in the human body, eyes also perceive light intensity logarithmically (see also Chapter XXX, as well as exercise X in Chapter XXX). This is eyes also obey Weber-Fechner law.

In order to perceive colors, the retina contains two major types of light-sensitive photoreceptors: the rods and the cones. The rods are responsible for monochrome perception and therefore enable black and white distinction. They are very sensitive, especially in low-light conditions. This is why darker scenes become more and more colorless. The cones are responsible for color vision. They require brighter light than the rods. In humans, there are three types of cones, maximally sensitive to long-wavelength, medium-wavelength, and short-wavelength light. The color perceived is the combined effect of stimuli to these three types of cone cells. Overall there are more rods than cones, so color perception is less accurate than black and white contrast perception. This affects the variety of perceived colors in contrast to gray tones as well as the accuracy of spatial color distinction in contrast to black and white. In other words, reducing the spatial resolution of the color representation while maintaining the black and white resolution has little perceptible effect. This property of the eye has been used heavily for compression, the analog TV format NTSC for example uses less bandwidth for color transmission than for black and white transmission. The JPEG image compression uses the spatial and the variance color insensitivity in multiple ways, as described in Chapter XXX.

Other leverageable properties of human vision are not based on anatomical properties of the eye but on functional properties of the brain behind it. These can become very complex and are most popularly studied in optical illusions. Some, if not most, of these properties are learned. For example, drawing a dark border on the lower and right edge of a window makes it appear in front of other windows because humans have learned to interpret the dark edge as shadow. There is evidence that even binocular depth perception is learned (see references).

Color Spaces

The authors found at point is this the right time to introduce some math. As discussed earlier, colors can be captured and reproduced by varying intensities of fixed colors. This concept is known from water color painting in elementary school: Red, blue, and yellow can be used in varying intensities to mix all the other colors. The CRT display uses red, green, and blue and the human eye uses yet a different set of filters based on pigmenting. Mathematically speaking, all colors can be described by a linear combination of base colors. In other words, the base colors from the basis of a 3-dimensional color space. Color spaces are a very important concept as different sensors and light reproducers can only work with a different set of fixed colors. Most printers use the CMYK (Cyan, Magenta, Yellow, BlacK) color space because it is most convenient for ink producers, the forth "K" stands for black, which in principle could be created by mixing yellow, magenta, and cyan -- however, this would be very costly. So even though "K" is mathematically not needed economic reasons prevail. Most importantly, color spaces are often

used to analyze an image or video computationally. In the following we present three important color spaces. Other color spaces will be introduced in further chapters of the book.

For computer scientists, the RGB color space is probably the canonical color space as most displays, graphics cards and raw image formats support this space. As a result, most programming tools, especially those for graphical user interfaces, work in this space by default. The RGB space is often augmented by a forth component, often called alpha, that controls the transparency of a pixel. It is important to know that the RGB color space is furthest away from human perception since contrast is not modeled explicitly. So the perceptual importance of a color component and the similarity of two colors cannot be judged easily.

For image compression, the YUV color space (and technical variants) has therefore used most predominantly. The YUV model defines a color space based on one luma (Y) and two so-called chrominance (UV) components. The YUV color model is used in the PAL and SECAM standards. Previous black-and-white systems used only luma information. Color information was added separately via a sub carrier signal so that a black-and-white receiver would still be able to receive and display a color picture transmission in the receiver's native black-and-white format. A variant of YUV is used for JPEG compression as it allows to scale color and black-and-white components independently. Conversion between RGB and YUV (and back) can be performed by a simple linear transformation:

$$\begin{bmatrix} Y' \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.14713 & -0.28886 & 0.436 \\ 0.615 & -0.51499 & -0.10001 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.13983 \\ 1 & -0.39465 & -0.58060 \\ 1 & 2.03211 & 0 \end{bmatrix} \begin{bmatrix} Y' \\ U \\ V \end{bmatrix}$$

The Y component is denoted with a prime symbol Y’ to indicate gamma adjustment of the Y component. A close look at the formula reveals the weighting of the different components which corresponds to experimental evidence for human color perception. Figure 8 shows the result of this decomposition for an example image.

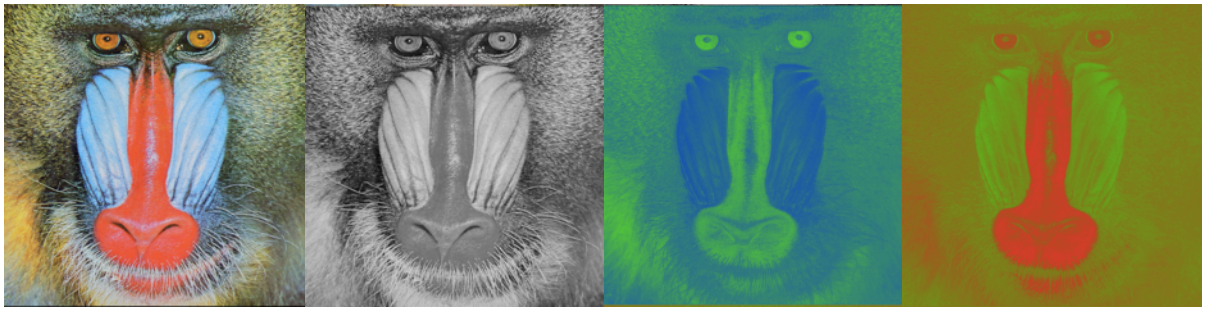

Figure 8. An image and it's Y, U, and V decompositions.

As the reader probably already suspects, a linear transformation from a color space that was invented for CRT displays is not able to describe human color perception exactly enough to measure color differences. Unfortunately, while extremely important, the measurement of perceived color differences is extremely difficult as perception of color differences varies not only with lighting but also with colors that surround the color difference. Also, there is a good deal of optical illusions that creates "fake" colors, i.e. colors that are not there but are perceived nevertheless (see references). Obviously, the objective color difference would be zero but the perceived color difference is greater than zero. Nevertheless, there is one color space which has been created to model perceived color differences using an abundancy of human-subject experiments. Also, it has recently gained attention in the computer vision and image retrieval communities. It is called the CIE LAB color space and is designed to be perceptually uniform, i.e. ideally the Euclidean distance between two colors reveals its perceptual difference. The CIELAB space is based on the opponent-colors theory of color vision. The theory assumes that two colors cannot be perceived as both green and red or blue and yellow at the same time. As a result, single values can be used to describe the red/green and the yellow/blue attributes. When a color is expressed in CIELAB, L defines lightness, a denotes the red/green value and b the yellow/blue value. Different standard illumination conditions are defined using a reference white. The most commonly used reference white is the so-called D65 reference white. CIELAB's perceptual color metric is still not optimal and the aforementioned assumption sometimes leads to problems. But in practice, the Euclidean distance between two colors in this space better approximates a perceptually uniform measure for color differences than in any other color space, like YUV or RGB. The color space uses an intermediate space, the so-called CIE XYZ sapce. The XYZ space was designed to eliminate metamers, i.e. different colors that are perceived as the same color. Figure 9 shows the color matching function used by the XYZ space.

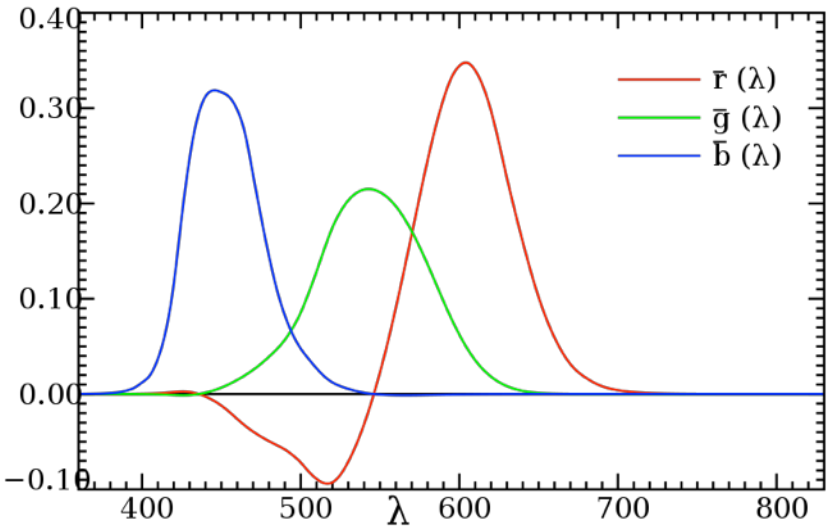

Figure 9. CIE XYZ space color matching function. The curves show the amount of primary color-mix needed to match the same monochromatic color generated by light at wavelength lambda.

The following formula converts RGB to CIE XYZ space:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Conversion from CIEXYZ to CIELAB is performed by the following formula:

$$\begin{aligned} L^\star &= 116 f(Y/Y_n) - 16 \\ a^\star &= 500 \left[f(X/X_n) - f(Y/Y_n) \right] \\ b^\star &= 200 \left[f(Y/Y_n) - f(Z/Z_n) \right] \end{aligned}$$

where

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases}$$

This chapter only provides only a quick introduction to light as relevant for multimedia computing. Further properties of light, human perception, and the devices that record and reproduce light will be discussed when important in connection to concrete algorithms further in this book.

Exercises

1. List the factors that contribute to an object reflecting more light than another one.
2. When a powerful light source and a not so powerful light source are placed adjacent to each other, the less powerful light source is sometimes appears to not even emit light (e.g. a small LED in the mid-day sun). Explain.
3. Write a program that can correct lens deformations using a calibration process, i.e. the photographed shape is known and a function is fitted to correct the photograph to the actual shape.
4. How much space is needed to store a raw image in NTSC and full HDTV format?
5. Take a pencil and hold it in front of your eyes. Close one eye, observe, open it again, then close the other eye and observe again. Repeat the experiment with the pencil at different distance in front of your eyes. What can you observe?
6. Describe a procedure to calibrate a 3D display with anaglyph technology and with parallax barrier technology.
7. How many bits are needed to store a pixel in CIELAB space?
8. Explain which part of visual perception is most often utilized by magicians doing magic tricks.
9. What is the equivalent of sound synthesis in the visual domain? What is the main issue when doing this?

Literature

- Wyzecki, G.; Stiles, W.S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae* (2nd ed. ed.). Wiley-Interscience.
- CIE. (1932). *Commission Internationale de l'Éclairage Proceedings, 1931*. Cambridge: Cambridge University Press.
- Guild, J. (1931). The colorimetric properties of the spectrum. Philosophical Transactions of the Royal Society of London, A230, 149-187.
- Stiles, W. S. & Burch, J. M. (1955). Interim report to the Commission Internationale de l'Éclairage Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching. Optica Acta, 2, 168-181.
- National Television System Committee (1951–1953), [Report and Reports of Panel No. 11, 11-A, 12-19, with Some supplementary references cited in the Reports, and the Petition for adoption of transmission standards for color television before the Federal Communications Commission, n.p., 1953], 17 v. illus., diagrs., tables. 28 cm. LC Control No.:54021386.
- *ITU-R BT.470-6, Conventional Television Systems*
- ITU-R Recommendation BT.709, *High-definition Television*

Web Links

- Atlas of Visual Phenomena: http://lite.bu.edu/vision/applets/lite/lite/lite.html

- Silencing: http://visionlab.harvard.edu/silencing/

Research Papers

- Crane, R. J. (1979). The Politics of International Standards: France and the Color TV War, Ablex Publishing Corporation.
- Fausto Bernardini, Holly E. Rushmeier (2002). "The 3D Model Acquisition Pipeline". *Comput. Graph. Forum* 21 (2): 149–172
- Brian Curless (November 2000). "From Range Scans to 3D Models". *ACM SIGGRAPH Computer Graphics* 33 (4): 38–41
- Song Zhang, Peisen Huang (2006). "High-resolution, real-time 3-D shape measurement" (PDF). *Optical Engineering*: 123601. http://www.math.harvard.edu/~songzhang/papers/Realtime_OE.pdf.
- Gonzalez, F. and Perez, R., Neural mechanisms underlying stereoscopic vision, Prog Neurobiol, 55(3), 191-224, 1998.
- Qian, N., Binocular Disparity and the Perception of Depth, Neuron, 18, 359-368, 1997.
- Zitnick, C. L. and Kanade, T. (2000). A Coopera- tive Algorithm for Stereo Matching and Occlusion Detection. IEEE Trans- actions on Pattern Analysis and Machine Intelligence, 22(7):675–684.
- A Time-Of-Flight Depth Sensor – System Description, Issues and Solutions. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington D.C., USA.
- Santrac, N., Friedland, G., and Rojas, R. (2006). High Resolution Segmentation with a Time-of-Flight 3D-Camera using the Exam- ple of a Lecture Scene. Technical Report B-06-09, Freie Universit ¨at Berlin, Institut fuer Informatik, Berlin, Germany.